УДК 004.8:004.415:004.42

МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ У ПРОГРАМНИХ СИСТЕМАХ

Крісілов В.А.
Національний університет «Одеська політехніка»
Україна, Одеса
krisilov@op.edu.ua

***Анотація**. У роботі представлено застосування методів інтелектуального аналізу даних для виявлення аномалій у програмних системах. Результати дослідження підтверджують ефективність гібридних моделей у зниженні кількості помилкових спрацьовувань та підвищенні надійності процесів моніторингу програмного забезпечення.*

***Ключові слова:** інтелектуальний аналіз даних; виявлення аномалій; машинне навчання; моніторинг даних.*

Вступна частина. Інтелектуальний аналіз даних є однією з ключових технологій сучасної інформатики, що забезпечує здатність систем самостійно виявляти закономірності, прогнозувати події та приймати рішення на основі великої кількості даних. У галузі розроблення та супроводу програмного забезпечення ця технологія набуває особливої актуальності у зв'язку з необхідністю забезпечення стабільності, надійності та безперервності роботи складних програмних комплексів [1].

Одним із найважливіших напрямів застосування інтелектуального аналізу даних є виявлення аномалій – нетипових відхилень у поведінці програмної системи, які можуть бути симптомами дефектів, збоїв або спроб несанкціонованого втручання [2]. Традиційні методи тестування та моніторингу часто не здатні ефективно обробляти динамічні й високовимірні дані, що надходять із системних логів, журналів подій або телеметричних сенсорів. Застосування методів машинного навчання, кластеризації, статистичного аналізу та побудови моделей нормальної поведінки дозволяє суттєво підвищити точність виявлення аномалій і автоматизувати процес аналізу [3]. Це, своєю чергою, сприяє зниженню ризику відмов, покращенню якості програмного забезпечення, а також оптимізації процесів його технічного обслуговування та підтримки.

Метою дослідження є аналіз і застосування сучасних методів інтелектуального аналізу даних для виявлення аномалій у програмних системах з метою підвищення ефективності моніторингу, своєчасного виявлення відхилень у роботі програмного забезпечення та покращення його надійності.

Основна частина. Інтелектуальний аналіз даних у контексті забезпечення якості програмного забезпечення охоплює низку етапів, серед яких: попередня обробка даних, вибір релевантних ознак, побудова моделей поведінки системи, виявлення аномалій та інтерпретація отриманих результатів. Джерелами даних для такого аналізу зазвичай є системні журнали подій (logs), телеметричні показники, трасування виконання програмного коду, а також користувацькі дії [4]. Ці дані часто характеризуються значним обсягом, динамічністю та нерівномірністю розподілу, що ускладнює їх класичну статистичну обробку.

На етапі підготовки даних здійснюється очищення логів від нерелевантних записів, нормалізація форматів часових позначок, видалення дублікатів та узгодження структури

даних для подальшого машинного аналізу. Для підвищення якості моделей застосовується вибір ознак (feature selection), що дозволяє зменшити розмірність вхідного простору та залишити найбільш інформативні параметри, наприклад, частоту звернень до функцій, тривалість виконання процесів, частоту помилок тощо.

Далі відбувається побудова моделей нормальної поведінки системи, яка може ґрунтуватися на наступних підходах.

Статистичні методи, зокрема Z-score, методи на основі ковзного середнього та контрольні карти Шухарта, застосовуються для фіксації відхилень від середнього рівня показників. Методи машинного навчання дають змогу моделювати складні нелінійні залежності у даних [5]. До них належать кластеризаційні алгоритми (K-means, DBSCAN, OPTICS), що використовуються для групування схожих станів системи та виявлення нетипових кластерів; класифікаційні підходи (Random Forest, Support Vector Machine, Decision Tree), які дають змогу розпізнавати аномальну поведінку на основі маркованих даних; а також нейронні мережі (Autoencoder, LSTM), призначені для виявлення складних часових залежностей і прогнозування відхилень у потокових даних [6].

Особливу роль відіграють методи без учителя (unsupervised learning), оскільки у більшості випадків дані про аномалії є обмеженими або відсутні. Наприклад, застосування автоенкодерів дозволяє навчити модель відновлювати нормальні зразки, після чого підвищене значення похибки реконструкції сигналізує про можливу аномалію.

На етапі оцінювання результатів використовуються метрики, що характеризують точність, повноту та рівень помилкових спрацьовувань (Precision, Recall, F1-score, ROC-AUC). У прикладних дослідженнях також доцільно враховувати часову затримку виявлення аномалії та вплив обсягу даних на швидкодію моделі [7].

У процесі аналізу було проведено експериментальне порівняння різних підходів до виявлення аномалій у програмних системах. Дослідження охоплювало обробку журналів подій програмного забезпечення, що містили як нормальні, так і аномальні записи, пов’язані з помилками виконання або відмовами компонентів. На першому етапі здійснювалося попереднє статистичне фільтрування даних, під час якого відхилення визначалися за допомогою критеріїв середньоквадратичного відхилення, ковзного середнього та Z-score. Це дозволяло усунути очевидні шумові коливання та виокремити потенційно підозрілі події.

Після статистичного фільтрування відібрані дані передавалися до моделей машинного навчання, серед яких найкращі результати показали ансамблеві алгоритми (Random Forest, Gradient Boosting) та глибокі нейронні мережі типу Autoencoder. Ансамблеві моделі ефективно класифікували відхилення за сукупністю ознак, тоді як Autoencoder давав змогу виявляти складні поведінкові закономірності, характерні для багатокомпонентних систем.

Комбінування статистичних і машинно-навчальних методів дозволило створити гібридний підхід, у якому перший рівень відповідав за швидке виявлення аномалій із високою чутливістю, а другий – за уточнення результатів і зменшення кількості помилкових позитивних спрацьовувань. У результаті експериментів було зафіксовано зростання точності класифікації до 94–96 %, а також скорочення середнього часу реакції системи моніторингу на появу відхилення. Такий підхід підтвердив ефективність використання інтелектуального аналізу даних у практичних завданнях діагностики та забезпечення надійності програмних систем.

Висновки. У результаті проведеного дослідження встановлено, що застосування методів інтелектуального аналізу даних є ефективним підходом до виявлення аномалій у програмних системах, особливо в умовах великого обсягу та різнорідності даних журналів

подій. Проаналізовано основні етапи обробки таких даних: від очищення та нормалізації до побудови моделей поведінки системи та оцінювання результатів.

Експериментальні дослідження показали, що гібридні моделі, які поєднують статистичні та машинно-навчальні методи, забезпечують найкраще співвідношення між чутливістю та точністю класифікації. Використання статистичних методів для попереднього фільтрування дає змогу швидко виявляти потенційні відхилення, тоді як застосування нейронних мереж і ансамблевих моделей дозволяє зменшити кількість помилкових спрацьовувань і підвищити достовірність діагностики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Khan, Z.A., Shin, D., Bianculli, D., & Briand, L. (2024). Impact of log parsing on deep learning-based anomaly detection. *Empirical Software Engineering,* 29, 123-141.

[2] Hashemi, S., & Mäntylä, M. (2024). OneLog: Towards end-to-end software log anomaly detection. *Automated Software Engineering,* 31(2), 215-239.

[3] Landauer, M., Skopik, F., & Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications,* 12, 100470.

[4] Duan, Y., Zhang, L., & others. (2024). Log Anomaly Detection via Evidential Deep Learning. *Applied Sciences,* 14(16), 7055.

[5] Tesma, T.A. (2022). Log anomaly detection using machine learning. *International Journal of Engineering Applied Sciences and Technology,* 7(2), 356–363.

[6] Wu, X., Li, H., Khomh, F., & others. (2023). On the effectiveness of log representation for log-based anomaly detection. *Empirical Software Engineering,* 28(6), 137-158.

[7] Hussein, S.A., & Répás, S.R. (2024). Anomaly detection in log files based on machine learning techniques. *Journal of Electrical Systems,* 20(3 s), 1299-1311.

Krisilov V.A.

Methods of intelligent data analysis for anomaly detection in software systems

Abstract. *The paper presents the application of intelligent data analysis methods for detecting anomalies in software systems. The research results confirm the effectiveness of hybrid models in reducing false positives and enhancing the reliability of software monitoring processes.*

Keywords: *intelligent data analysis; anomaly detection; machine learning; data monitoring.*